

Facility Location in Evolving Metrics*

David Eisenstat[†] Claire Mathieu[‡] Nicolas Schabanel[§]

March 27, 2014

Abstract

Understanding the dynamics of evolving social or infrastructure networks is a challenge in applied areas such as epidemiology, viral marketing, or urban planning. During the past decade, data has been collected on such networks but has yet to be fully analyzed. We propose to use information on the dynamics of the data to find stable partitions of the network into groups. For that purpose, we introduce a time-dependent, dynamic version of the facility location problem, that includes a switching cost when a client's assignment changes from one facility to another. This might provide a better representation of an evolving network, emphasizing the abrupt change of relationships between subjects rather than the continuous evolution of the underlying network. We show that in realistic examples this model yields indeed better fitting solutions than optimizing every snapshot independently. We present an $O(\log nT)$ -approximation algorithm and a matching hardness result, where n is the number of clients and T the number of time steps. We also give an other algorithms with approximation ratio $O(\log nT)$ for the variant where one pays at each time step (leasing) for each open facility.

1 Introduction

During the past decade, a massive amount of data has been collected on diverse networks such as web links, nation- or world-wide social networks, online social networks (Facebook or Twitter for example), social encounters in hospitals, schools, companies, or conferences (e.g. [18, 21]), and other real-life networks. Those networks evolve with time, and their dynamics have a considerable impact on their structure and effectiveness (e.g. [19, 14]). Understanding the dynamics of evolving networks is a central question in many applied areas

*This work was partially supported by the ANR-2010-BLAN-0204 Magnum and ANR-12-BS02-005 RDM grants.

[†]Brown University (USA)

[‡]CNRS, École normale supérieure UMR 8548 (France) - <http://www.di.ens.fr/ClaireMathieu.html>

[§]CNRS, Université Paris Diderot (France) - <http://www.liafa.univ-paris-diderot.fr/~nschaban/>
- IXXI, École normale supérieure de Lyon (France)

such as epidemiology, vaccination planning, anti-virus design, management of human resources, viral marketing, “facebooking”, etc. Obtaining a relevant clustering of the data is often a key to the design informative representations of massive data sets. Algorithmic approaches have for instance been successful in yielding useful insights on several real networks such as zebras social interaction networks [22].

But the dynamics of real-life evolving networks are not yet well understood, partly because it is difficult to observe and analyze such large networks sparsely connected over time. Some basic facts have been observed (such as the preferential attachment or copy-paste mechanisms) but more specific structures remain to be discovered. In this article, we propose a new formulation of the facility location problem adapted to these evolving networks. We show that requiring the solution to be stable over time yields in many realistic situations better fitting solutions than optimizing independently various snapshots of the network.

The problem. We focus on the facility location problem where clients are moving in some space: we look for the best connections of clients to facilities (sometimes called centers) over time minimizing a tradeoff between three objectives. The two first objectives are classically: the *distance cost* of the connections (the sum of the connection lengths), so that each client gets connected to a facility representative of its position; and the *opening cost*, a price paid for opening each open facility, so that only the most meaningful facilities (and as few of them as possible) get open. The third and new objective is the *instability* of the connections over time, measured as the number of clients switching from one facility to another over time, so that only the changes responding to significative and lasting changes in the metrics get authorized. We argue that incorporating this stability requirement in the objective function helps in many realistic situations to obtain more desirable solutions (see Section 2.1).

Related work. Facility location problem has been studied extensively in the offline, online and incremental settings, see [12] for a survey. The offline version of the problem was a case study accompanying the development of techniques for approximation algorithms: primal-dual and dual fitting methods and local search for example. A series of papers, [20, 16, 13, 2, 5, 3, 15], obtained (almost) matching upper and lower bounds on the polynomially achievable approximation ratio in this setting: $\Theta(\log n)$ in the non-metric case, and within $[1.463, 1.488]$ in the metric case, when the client-to-facility connection cost is a distance in a metric space.

The online setting, where clients arrive over time and the algorithm gradually buys more and more facilities to serve them, was first addressed by [17] who obtained $\Theta(\log n / \log \log n)$ upper and lower bounds on the competitive ratio of any online algorithm. This later led to developments for various cases, e.g. analyzed when clients are drawn from some distribution [1] and in other cases [11]. In order to allow more flexibility in the solution (as required in many clustering application), incremental approaches, which allow reconsidering the assignment of clients to facilities over time, were also considered. Such variants may allow better ($O(1)$) competitive ratios, see e.g. in the metric case but with streaming constraints [10], and in the special case in the Euclidian setting when facilities may be moved as new clients

arrive [8]. We also mention the related clustering problem in which clusters may be merged but not split, e.g. [4].

Our approach differs from the existing algorithmic approaches to dynamic settings because we focus on settings where distances may vary over time, and where it is desirable to achieve a tradeoff between the *stability* of the solution — clusters of clients tend not to flip-flop constantly — and its *adaptability* — the assignment ought to be modified if distance change too much. We show that offline static algorithms that construct an independent optimal solution for each snapshot of the network yield results that, in a large variety of realistic situations, are not only unstable (and thus arbitrarily bad for our objective), but also undesirable with respect to network dynamics analysis. Online solutions such as the clustering of Charikar et al. [4] are also unnecessarily pessimistic in this setting: we have access to the whole evolution of the network over time (as given by experiments such as [21]) and we can thus anticipate future changes.

As far as we know, the case where the distance between points vary overtime is still largely unexplored.

Our results. After defining the problem formally in section 2.1 and giving examples showing the benefits one can expect from solving this problem in the context of metrics evolving with time, we give in Section 2.2 a $O(\log nT)$ -approximation algorithm for this problem, where n is the number of clients and T the number of time steps.

Theorem 1 (Fixed opening cost). *There is a polynomial time randomized algorithm which outputs a solution to the dynamic facility location problem with fixed opening cost whose cost verifies:*

$$\Pr\{\text{cost} \leq 4 \log(2nT) \cdot \text{OPT}\} \geq \Pr\{\text{cost} \leq 4 \log(2nT) \cdot \text{LP}\} \geq 1/4.$$

Repeating this algorithm $O(\log \frac{1}{\epsilon})$ times and outputting the best solution increases the success probability to $1 - \epsilon$ for arbitrarily small $\epsilon > 0$. We then show in Section 2.3 that this approximation ratio is asymptotically optimal, even if one assume that the distance verifies the triangle inequality at every time step and if the input consists in only one client and two possible positions for the client and the facilities.

Theorem 2 (Hardness for fixed opening cost). *Unless $P \neq NP$, there is no $o(\log T)$ -approximation, even for the metric case with one client and two possible positions.*

This new problem differs then significantly from the classic facility location problem which admits no $o(\log n)$ -approximation for non-metric distances but a 1.488-approximation when the distance satisfy the triangle inequality [15]. We then show in Section 3 how to extend our approximation algorithm to the setting where facilities can be open and closed at any time step and where one pays f for each facility open at each time step.

Theorem 3 (Hourly opening cost). *There is a polynomial time randomized algorithm which outputs a solution to the dynamic facility location problem with hourly opening cost whose cost verifies:*

$$\Pr\{\text{cost} \leq 4 \log(2nT) \cdot \text{OPT}\} \geq \Pr\{\text{cost} \leq 4 \log(2nT) \cdot \text{LP}\} \geq 1/4.$$

This article concludes with several open questions and possible extension of this work.

2 Facility Location in Evolving Metrics

2.1 Definition

Dynamic Facility Location problem with fixed opening cost. We are given a set F of m *facilities* and a set C of n *clients* together with a finite sequence of distances $(d_t)_{1 \leq t \leq T}$ over $F \times C$, a non-negative *facility opening cost* f and a non-negative *client switching cost* g . The goal is to output a subset $A \subseteq F$ of facilities and, for each time step $t \in [T]$, an assignment $\phi_t : C \rightarrow A$ of facilities to clients, so as to minimize:

$$f \cdot \#A + \sum_{1 \leq t \leq T, j \in C} d_t(\phi_t(j), j) + g \cdot \sum_{1 \leq t < T} \sum_{j \in C} \mathbb{1}\{\phi_t(j) \neq \phi_{t+1}(j)\},$$

that is to say the sum of the opening cost (f for each open facility), of the total *distance cost* to connect each client to its assigned facility at every time step, and of the *switching cost* for each client (g per change of facility per client).

Examples. The two examples in Figure 1 show how facility location in the dynamic setting is quite different from the static setting and yields more desirable partitions of the clients. In both examples, a facility can be opened at every client (so that electing a facility consists in electing a representative for every significantly different behavior).

In example 1(a), we see a classroom with students split into five groups and a teacher moving from group to group in cyclic order. When the number of students is large, static facility location isolates the five groups and moves the teacher from one group to the next between snapshots; whereas dynamic facility location isolates every group of students and puts the teacher in a sixth group.

In example 1(b) we see two groups of people crossing each other (on a street for instance): a static facility location would first output the two groups, then merge them into a single group, then split it into two groups again; whereas a dynamic facility location would keep the same groups for the whole time period, with the same representatives.

Assuming in both examples that the distances between individuals are either very small or very large, then the ratio of the (dynamic) cost between the dynamic solution and the sequence of static solutions can be made arbitrary large because the switching cost grows for the sequence of static solutions as $\Omega(T)$ and $\Omega(n)$ respectively.

Fact 4. *The ratio between the cost of an optimal dynamic facility location solution and the (dynamic) cost of a sequence of optimal static facility location solutions for each snapshot can be as large as $\Omega(T)$ and $\Omega(n)$.*

A linear relaxation. For an integer programming formulation, we define indicator 0-1 variables y_i , x_{ij}^t , and z_{ij}^t for $i \in F$, $j \in C$, and $t \in [T]$: $y_i = 1$ iff facility i is open; $x_{ij}^t = 1$ iff client j is connected to facility i at time t ; and $z_{ij}^t = 1$ iff client j is connected to facility i at time t but no more at time $t + 1$. The dynamic facility location problem is then equivalent to finding an integer solution to the following linear programming relaxation.

$$\left\{ \begin{array}{l} \text{Minimize} \quad f \cdot \sum_{i \in F} y_i + \sum_{j \in C} \sum_{1 \leq t \leq T} \sum_{i \in F} x_{ij}^t \cdot d_t(i, j) + g \cdot \sum_{j \in C} \sum_{1 \leq t < T} \sum_{i \in F} z_{ij}^t \\ \text{such that:} \quad (\forall i, j, t) \quad x_{ij}^t \leq y_i \\ \quad \quad \quad (\forall j, t) \quad \sum_{i \in F} x_{ij}^t = 1 \\ \quad \quad \quad (\forall i, j, \forall t < T) \quad z_{ij}^t \geq x_{ij}^t - x_{ij}^{t+1} \\ \quad \quad \quad y_i, x_{ij}^t, z_{ij}^t \geq 0 \end{array} \right. \quad (1)$$

2.2 Approximation algorithm

Algorithm 1 Fixed opening cost

- Solve the linear program LP (1). Let (x, y, z) be the solution obtained.
 - Draw a facility at random $\Gamma = 2 \log(2nT) \sum_{i \in F} y_i$ times independently, with distribution proportional to y ; let A be the resulting multiset of facilities.
- for** For each client j **do**
- Determine when it should change from one facility to another using the z -variables, and assign it to the cheapest selected facility between each change:
 - (a) Partition time greedily into ℓ_j intervals $[t_k^j, t_{k+1}^j)$ such that $t_1^j = 1$ and where t_{k+1}^j is inductively defined as the largest $t \in (t_k^j, T + 1]$ such that $\sum_{i \in F} \left(\min_{t_k^j \leq u < t} x_{ij}^u \right) \geq 1/2$, and $t_{\ell_j+1}^j = T + 1$;
 - (b) For each time interval $[t_k^j, t_{k+1}^j)$, connect j to the facility in A that is cheapest for j for that time interval.
- end for**
-

Theorem 1 (page 3) states that Algorithm 1 outputs a $O(\log nT)$ -approximation with positive constant probability. The next section will show that this is asymptotically optimal (unless $P \neq NP$).

Proof of Theorem 1. The expected facility opening cost is obviously at most $2f \log(2nT) \sum_{i \in F} y_i$. In order to bound the switching and distance costs, let us now fix a client j and show the following fact:

Fact 5. For all clients j and time intervals $[t_k^j, t_{k+1}^j)$ with $k < \ell_j$:

$$\sum_{t_k^j \leq t < t_{k+1}^j} \sum_{i \in F} z_{ij}^t > 1/2.$$

This fact yields an easy bound on the switching cost: the switching cost for client j is g times the number of its intervals minus 1. But according to the fact above, for every interval except the last one, the z_{ij}^t 's sum to at least $1/2$, so LP (1) pays at least $g/2$ for that interval. The switching cost of the solution is then at most twice the corresponding term in the LP.

Proof of Fact 5. For all t , $\sum_{i \in F} x_{ij}^t = 1$, in particular at time t_k^j . Now, since $k < \ell_j$, we have $\sum_{i \in F} \left(\min_{t_k^j \leq t \leq t_{k+1}^j} x_{ij}^t \right) < 1/2$. Let $t_i \in [t_k^j, t_{k+1}^j]$ such that $x_{ij}^{t_i} = \min_{t_k^j \leq t \leq t_{k+1}^j} x_{ij}^t$. We have $\sum_{i \in F} x_{ij}^{t_i} < 1/2$. Now, as $z_{ij}^t \geq 0$, $\sum_{t_k^j \leq t < t_{k+1}^j} z_{ij}^t \geq \sum_{t_k^j \leq t < t_i} z_{ij}^t \geq \sum_{t_k^j \leq t < t_i} (x_{ij}^t - x_{ij}^{t+1}) = x_{ij}^{t_k^j} - x_{ij}^{t_i}$. It follows that $\sum_{t_k^j \leq t < t_{k+1}^j} \sum_{i \in F} z_{ij}^t \geq \sum_{i \in F} x_{ij}^{t_k^j} - \sum_{i \in F} x_{ij}^{t_i} > 1 - 1/2 = 1/2$. \square

Let us now bound the expected distance cost for client j within each interval $I = [t_k^j, t_{k+1}^j]$. Let $x_{ij}^I = \min_{t \in I} x_{ij}^t$ and $\hat{x}_{ij}^I = x_{ij}^I / \sum_{i \in A} x_{ij}^I$. We want to argue that the facility selection process (which is according to the y_i 's) can be simulated, to within a factor of 2, by selecting a facility according to \hat{x}_{ij}^I . Then the expected distance is correct up to a factor of 2.

We know that $x_{ij}^t \leq y_i$. We can view sampling proportionally to (y_i) as: with probability $p_j^I = \frac{\sum_i x_{ij}^I}{\sum_i y_i}$, sample proportionally to x_{ij}^I , and with the remaining probability, sample proportionally to $y_i - x_{ij}^I$. Indeed,

$$\begin{aligned} \Pr\{i \text{ is selected by this process}\} &= \frac{\sum_i x_{ij}^I}{\sum_i y_i} \cdot \frac{x_{ij}^I}{\sum_i x_{ij}^I} + \left(1 - \frac{\sum_i x_{ij}^I}{\sum_i y_i}\right) \cdot \frac{y_i - x_{ij}^I}{\sum_i (y_i - x_{ij}^I)} \\ &= \frac{x_{ij}^I}{\sum_i y_i} + \frac{\sum_i y_i - \sum_i x_{ij}^I}{\sum_i y_i} \cdot \frac{y_i - x_{ij}^I}{\sum_i y_i - \sum_i x_{ij}^I} \\ &= \frac{y_i}{\sum_i y_i}. \end{aligned}$$

Formally, we consider the following facility selection process: let U be a uniform real number in $[0, \sum_i y_i)$, we say that facility i is selected if $U \in [\sum_{k < i} y_k, \sum_{k \leq i} y_k)$ and that event B_j^I occurs if $U \in [\sum_{k < i} y_k, \sum_{k < i} y_k + x_{ij}^I)$. As pointed out before, according to this process: 1) i is distributed proportionally to y_i ; 2) $\Pr B_j^I = p_j^I = \frac{\sum_i x_{ij}^I}{\sum_i y_i} \geq \frac{1}{2 \sum_i y_i}$; and 3) conditioned to event B_j^I , i is distributed proportionally to x_{ij}^I .

We repeat the selection process $2 \log(2nT) \sum_i y_i$ times independently. Given a pair (j, I) , the probability that event B_j^I never occurs is at most $(1 - p_j^I)^{2 \log(nT) \sum_i y_i} \leq \exp\left(-\frac{2 \log(2nT) \sum_i y_i}{2 \sum_i y_i}\right) = \frac{1}{2nT}$. Since there are at most nT pairs (j, I) , the union bounds ensures that with probability at least $\frac{1}{2}$, all the events B_j^I occur at least once during the selection.

When B_j^I occurs, the facility i is selected according to x_{ij}^I . It follows that the expected distance of this selected facility to j is for all time $t \in I$: $\sum_i \frac{x_{ij}^I}{\sum_i x_{ij}^I} \cdot d_t(i, j) \leq \frac{1}{1/2} \sum_i x_{ij}^t d_t(j, i)$.

It follows that with probability at least $\frac{1}{2}$, the expectation of the sum of the distances of all j 's at all time t to their closest selected facility in F is at most: $2 \sum_{j,t} \sum_i x_{ij}^t d_t(j, i)$. Summing all the contribution, with probability at least $\frac{1}{2}$, the expected cost of the solution is at most:

$$2f \log(2nT) \sum_i y_i + 2 \sum_{i,j,t} x_{ij}^t d_t(j, i) + 2g \sum_{i,j,t} z_{ij}^t \leq 2 \log(2nT) \cdot LP.$$

Applying Markov inequality we get:

$$\begin{aligned} & \Pr\{cost \leq 4 \log(2nT) LP\} \\ & \geq \Pr\{cost \leq 2 \cdot 2 \log(2nT) LP \text{ and all } B_j^I \text{ occur}\} \\ & = \Pr\{cost \leq 2 \cdot 2 \log(2nT) LP \mid \text{all } B_j^I \text{ occur}\} \cdot \Pr\{\text{all } B_j^I \text{ occur}\} \\ & \geq \frac{1}{2} \cdot \frac{1}{2}. \quad (\text{Markov}) \end{aligned}$$

□

2.3 Hardness of approximation

Proof of Theorem 2. We do a reduction from Set Cover.

Pick an instance of set cover with T elements and m sets. We define the following instance of dynamic facility location. There is one timestep t for each element of the set cover instance, one facility i for each set of the set cover instance, and a single client. We set $g = 0$ (i.e., g is small enough w.r.t. f , $1/n$ and $1/T$). Assume the only possible locations for the client and facilities are two points a and b at distance ∞ (i.e. large enough) from each other (note that it satisfies the triangle inequality). At every time step, the client sits at location a . For each set i of the set cover instance, the corresponding facility is in position a if set i contains element t , and in position b otherwise. This defines the instance of dynamic facility location.

Since the distance is infinite between the two locations, the output A , to have finite cost, has to correspond to a cover of the unique client for all T time steps, i.e. a cover of all T elements in the set cover input. The cost is then simply f times the number of selected facilities. The $\Omega(\ln T)$ hardness lower bound for Set Cover with T elements in [7] implies that same lower bound for our problem. □

3 Hourly opening cost

3.1 Dynamic Facility Location with hourly opening cost

We now focus on a variant of the problem studied in the previous section where the facilities can be open and closed at any time step and where the opening cost f is paid for every facility open at every time step.

Dynamic Facility Location problem with *hourly* opening cost. We are given a set F of m facilities and a set C of n clients together with a finite sequence of distances $(d_t)_{1 \leq t \leq T}$ over $F \times C$, and two non-negative values f and g . The goal is to output a sequence of subsets $A_t \subseteq F$ of facilities and, for each time step $t \in [T]$ an assignment, $\phi_t : C \rightarrow A_t$ of facilities to clients, so as to minimize:

$$f \cdot \sum_{1 \leq t \leq T} \#A_t + \sum_{1 \leq t \leq T, j \in C} d_t(\phi_t(j), j) + g \cdot \sum_{1 \leq t < T} \sum_{j \in C} \mathbb{1}\{\phi_t(j) \neq \phi_{t+1}(j)\}.$$

Linear relaxation. The LP (1) readily extends, with variables y_i^t :

$$\left\{ \begin{array}{l} \text{Minimize} \quad f \sum_{1 \leq t \leq T} \sum_{i \in A} y_i^t + \sum_{j \in C} \sum_{1 \leq t \leq T} \sum_{i \in F} x_{ij}^t \cdot d_t(i, j) + g \sum_{j \in C} \sum_{1 \leq t < T} \sum_{i \in F} z_{ij}^t \\ \text{such that:} \quad (\forall i, j, t) \quad x_{ij}^t \leq y_i^t \\ \quad (\forall j, t) \quad \sum_{i \in F} x_{ij}^t = 1 \\ \quad (\forall i, j, \forall t < T) \quad z_{ij}^t \geq x_{ij}^t - x_{ij}^{t+1} \\ \quad y_i^t, x_{ij}^t, z_{ij}^t \geq 0 \end{array} \right. \quad (2)$$

3.2 $O(\log nT)$ -Approximation algorithm

We now change the sampling procedure for the facilities: every facility i selects an exponentially distributed random threshold and opens only when its y_i^t variable is above the threshold.

Algorithm 2 Hourly opening cost

- Solve the linear program LP (2). Let (x, y, z) be the solution obtained.
- For each facility i , pick a random threshold ρ_i according to an exponential distribution with expectation $1/(2 \log(2nT))$: i.e. $\Pr\{\rho_i > a\} = e^{-2a \log(2nT)}$ for all $a \geq 0$. Open facility i at all times t such that $y_i^t > \rho_i$. For each time t , let A_t be the resulting multiset of facilities open at t .

for each client j **do**

- Determine when it should change from one facility to another using the z -variables, and assign it to the cheapest selected facility between each change:

(a) As before, partition time greedily into ℓ_j intervals $[t_k^j, t_{k+1}^j)$ s.t. $t_1^j = 1$ and where t_{k+1}^j is inductively defined as the largest $t \in (t_k^j, T + 1]$ with $\sum_{i \in F} \left(\min_{t_k^j \leq u < t} x_{ij}^u \right) \geq 1/2$, and

$$t_{\ell_j+1}^j = T + 1;$$

(b) For each time interval $I = [t_k^j, t_{k+1}^j)$ and facility i , let $x_{ij}^I = \min_{t \in I} x_{ij}^t$ and connect client j to the facility $i \in A_t$ that minimizes the ratio ρ_i / x_{ij}^I .

end for

The idea is that client j selects a facility i whose opening threshold is below its x_{ij}^I -value. We will show that one can find such a facility for all clients at all time steps with high probability.

Let us now analyze the cost of the resulting solution (Theorem 3).

Lemma 6. *The expected opening cost is at most $2 \log(2nT)$ times the corresponding term in LP (2).*

Proof. Facility i is open at time t with probability $\Pr\{\rho_i \leq y_i^t\} = 1 - e^{-2y_i^t \log(2nT)} \leq y_i^t \log(2nT)$ since $e^a \geq 1 + a$ for all $a \in \mathbb{R}$. The expected facility cost is thus at most $f \sum_i \sum_t y_i^t \cdot 2 \log(2nT)$. \square

As before, Fact 5 holds here as well and the total switching cost is at most twice the value of the corresponding in LP (2).

Proof of Theorem 3. We are now left with evaluating the distance cost. We want to show that we can view things so that there is a facility sampled according to x_{ij}^I that is alive throughout the time interval I . Use the same ρ_i 's, but imagine that you only open facility i if $x_{ij}^I > \rho_i$. Since we assign client j to the facility i such that ρ_i/x_{ij}^I is minimum, we do get a facility in that way, as long as $\rho_i/x_{ij}^I < 1$. Note that ρ_i/x_{ij}^I is an exponential of rate $2x_{ij}^I \log(2nT)$, and by independence of the ρ_i 's, $\min_i(\rho_i/x_{ij}^I)$ is also an exponential of rate $2 \sum_i x_{ij}^I \log(2nT)$, indeed:

$$\Pr\{\min_i(\rho_i/x_{ij}^I) > a\} = \prod_i \Pr\{\rho_i > a \cdot x_{ij}^I\} = e^{-2a \cdot (\sum_i x_{ij}^I \log(2nT))}.$$

Then, the probability that client i is not covered by an open facility by this process during time interval I is $\Pr\{\min_i(\rho_i/x_{ij}^I) \geq 1\} = e^{-2(\sum_i x_{ij}^I) \log(2nT)} \leq \frac{1}{2nT}$ since $\sum_i x_{ij}^I \geq 1/2$. Consider now the event B that for all client j and all interval I , $\min_i(\rho_i/x_{ij}^I) < 1$, then $\Pr B \geq \frac{1}{2}$ by the union bound. Now, conditioned to event B , the expected distance between every client i to the facility it is assigned during each interval I is upper bounded by:

$$\begin{aligned} \sum_i \frac{\Pr\{\rho_i \leq x_{ij}^I\}}{\Pr B} d_t(i, j) &\leq \sum_i 2 \left(1 - e^{-2x_{ij}^I \log(2nT)}\right) d_t(i, j) \\ &\leq \sum_i 4x_{ij}^I \log(2nT) d_t(i, j) \end{aligned}$$

The expected distance cost conditioned to event B is thus at most $4 \log(2nT)$ times the corresponding term in LP (2).

As $\Pr B \geq \frac{1}{2}$, the expected facility cost conditioned to event B is at most twice the unconditioned facility cost. The expected overall cost conditioned to event B is then at most:

$$4 \log(2nT) f \sum_{i,t} y_i^t + 2g \sum_{i,j,t} z_{ij}^t + 4 \log(2nT) \sum_{i,j,t} x_{ij}^I d_t(i, j) \leq 4 \log(2nT) \text{ LP}$$

We conclude by applying Markov inequality as before. \square

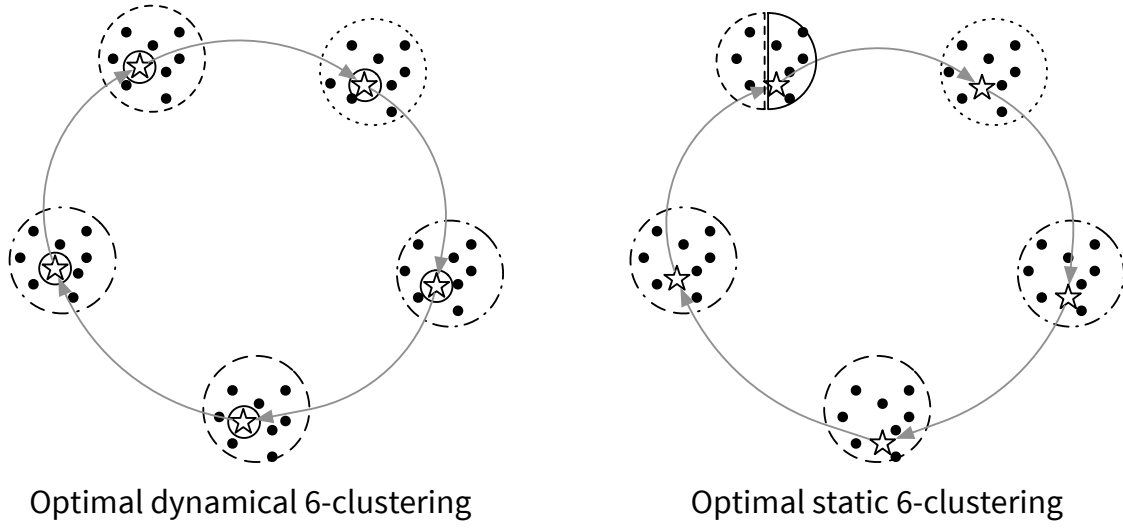
4 Conclusion and open questions

Algorithm 1 applies even if the distance do not follow the triangle inequality, and extends directly to non-uniform opening cost as well as to arrival and departures dates for clients. It is striking that instances with distances verifying the triangle inequality are not easier in the dynamic setting as opposed to the classic static setting (the approximation ratio $\Theta(\log nT)$ of Algorithm 1 is tight in both dynamic cases). Algorithm 2 extends also directly to the setting of opening costs which are non-uniform in time as well. The last section raises naturally the question whether there is an $\omega(1)$ -hardness result / $O(1)$ -approximation algorithm for the general hourly opening cost case. We believe that our dynamic setting should be helpful in designing better *static* representations of dynamical graphs (such as two dimensional flowcharts of the clients navigating between the different facilities over time). An other natural extension of our work is to study other objective functions for the distance cost, such as the sum of the diameters of resulting clusters at all time (i.e. the sum of the distance of the farthest client attached to each facility at all time, see e.g. [6] for a static formulation). As it turns out, the optimal dynamic solutions tend to adopt very intriguing behaviors under this objective, even in the simplest case of client moving along a fixed line, as has been observed in [9].

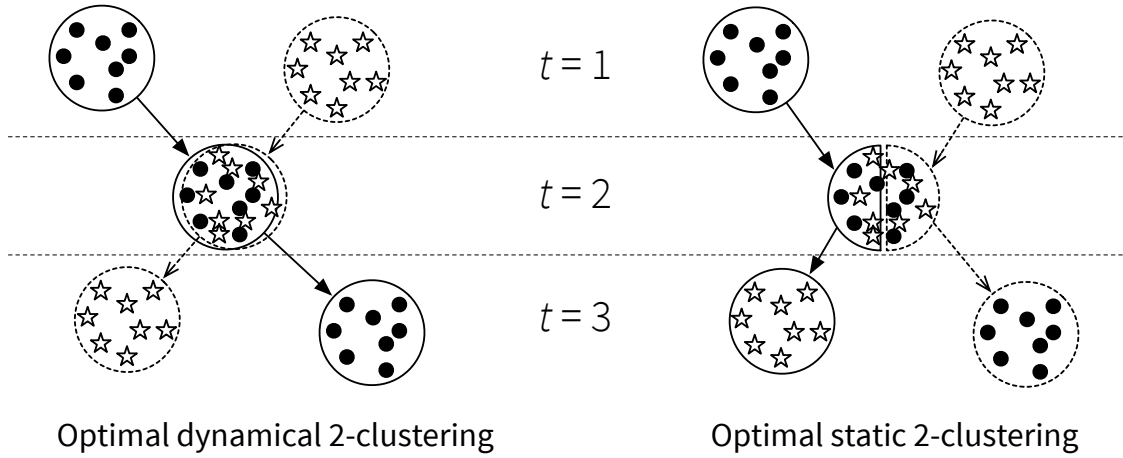
References

- [1] A. Anagnostopoulos, R. Bent, E. Upfal, and P. Van Hentenryck. A simple and deterministic competitive algorithm for online facility location. *Information and Computation*, 194:175–202, 2004.
- [2] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k -median and facility location problems. *SIAM J. on Computing*, 33(3):544–562, 2004.
- [3] J. Byrka and K. Aardal. An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem. *SIAM J. on Computing*, 39(6):2212–2231, 2010.
- [4] M. Charikar, C. Chekuri, T. Feder, and R. Motwani. Incremental clustering and dynamic information retrieval. In *STOC*, pages 626–635, 1997.
- [5] M. Charikar and S. Guha. Improved combinatorial algorithms for facility location problems. *SIAM J. on Computing*, 34(4):803–824, 2005.
- [6] M. Charikar and R. Panigrahy. Clustering to minimize the sum of cluster diameters. In *STOC*, pages 1–10, 2001.
- [7] Irit Dinur and David Steurer. Analytical approach to parallel repetition. In *STOC 2014*, *arXiv:1305.1979*, 2014.

- [8] G. Divéki and C. Imreh. Online facility location with facility movements. *Central European J. of Operations Research*, 19(2):191–200, 2010.
- [9] Cristina G. Fernandes, Marcio I. Oshiro, and Nicolas Schabanel. Dynamic clustering of evolving networks: some results on the line. In *AlgoTel*, 2013. Url: hal-00818985 (4 pages).
- [10] D. Fotakis. Incremental algorithms for facility location and k -median. *Theoretical Computer Science*, 361(2-3):275–313, 2006.
- [11] D. Fotakis. On the competitive ratio for online facility location. *Algorithmica*, 50(1):1–57, 2008.
- [12] D. Fotakis. Online and incremental algorithms for facility location. *SIGACT News*, 42(1):97–131, 2011.
- [13] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing lp. *J. ACM*, 50(6):795–824, 2003.
- [14] J. M. Kleinberg. The small-world phenomenon and decentralized search. *SIAM News*, 37(3), 2004.
- [15] Shi Li. A 1.488-approximation algorithm for the uncapacitated facility location problem. In *ICALP*, pages 77–88, 2011.
- [16] M. Mahdian, Y. Ye, and J. Zhang. Improved approximation algorithms for metric facility location problems. In *APPROX*, volume LNCS 2462, pages 229–242, 2002.
- [17] A. Meyerson. Online facility location. In *FOCS*, volume 42, pages 426–431, 2001.
- [18] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
- [19] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86:3200–3203, 2001.
- [20] D. B. Shmoys, E. Tardos, , and K. I. Aardal. Approximation algorithms for facility location problems. In *STOC*, volume 29, pages 265–274, 1997.
- [21] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina, and P. Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE*, 6(8):e23176, 2011.
- [22] C. Tantipathananandh, T. Y. Berger-Wolf, and D. Kempe. A framework for community identification in dynamic social networks. In *KDD*, pages 717–726, 2007.



(a) The classroom: one teacher cycling between 5 groups of students.



(b) Two groups crossing.

Figure 1: Dynamic versus static Facility Location.